A Novel Field Approach to 3D Gene Expression Pattern Characterization

L. da F. Costa,* B. Travençolo,† and A. Azeredo
Instituto de Física de São Carlos, Universidade de São Paulo,
Av. Trabalhador São Carlense 400, Caixa Postal 369, CEP 13560-970, São Carlos, SP, Brazil

M. E. Beletti
Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia,
Av. Pará 1720, CEP 38400-902, Uberlândia, MG, Brazil

D. Rasskin-Gutman, G. Sternik, J. C. I. Belmonte, and M. Ibañes
Salk Institute, 10010N Torrey Pines Road, La Jolla, USA CA 92037

G. B. Müller
Department of Zoology, University of Vienna, Althaustrasse 14, A-1090 Vienna Austria.
(Dated: February 9, 2008)

We present a vector field method for obtaining the spatial organization of 3D patterns of gene expression based on gradients and lines of force obtained by numerical integration. The convergence of these lines of force in local maxima are identified as centers of gene expression, providing a natural and powerful framework to characterize the organization and dynamics of biological structures. We apply this novel methodology to analyze the expression patterns of light chain myosin II protein linked to enhanced green fluorescent protein (EGFP) during zebrafish heart formation.

PACS numbers:

Animal development involves synchronized gene activation modulated by environmental influences [1, 2]. Far from being uniform, such a gene expression gives rise to structured spatial and temporal patterns of varying protein concentration. Recent advances in biochemical and imaging methods have paved the way to obtaining 3D reconstructions of spatial gene activation [3] which can be analyzed in order to better understand the intricate mechanisms governing tissue, organ and member formation [2]. Among the several currently available methodologies allowing characterization of 3D gene expression, special attention has been given to EGFP (Enhanced Green Fluorescence Protein). The EGFP is used as a marker. Its expression is controlled by the promoter of the gene of interest creating a fluorescent fusion protein that maintains the normal functions and localization of the wild type protein. This methodology can be used to demonstrate gene activity in intact cells and organisms, while taking into account the fact that the host protein is continuously synthesized, degraded, and suffering alterations within cells [4, 5]. As such a type of gene expression data becomes available, it is important to identify and develop mathematical methologies for measuring and modeling spatial gene activation. In addition to traditional approaches (e.g. density or dispersion estimation), it is important to consider more sophisticated methods capable of addressing more directly aspects related to the dynamics of the involved biological processes, such as cell communication and migration [6, 7], which play an important role during both embryonic development and pathological processes.

In this article we characterize the spatial organization of gene expression patterns in order to assess the geometrical basis of some dynamical processes during morphogenesis. To this end, we compute a "gene expression landscape" as a scalar field $\omega = g(x, y, z)$, where ω is interpreted as the amount of expression of the protein in the spatial position (x, y, z). The same approach can be used to model and predict the dissemination of cell signalling or other influence factors emmanating from the cell under analysis which, combined with the possibility of adopting varying values of the parameters affecting the field (e.g. the dielectric constant), defines a truly general framework for expressing field influences. In analogy with the potential dynamics of dissipative systems, we obtain the spatial trajectories (lines of force) corresponding to maximizing the gradient of gene expression. Such trajectories tend to converge to local peaks of activity, defining gene expression centers. It is proposed in this article that the distribution of such centers provide a natural framework for characterizing and analyzing the spatial interactions between the involved developmental rudiments. The potential of such a methodology is illustrated with respect to the analysis of zebrafish heart formation from 3D gene expression data.

Zebrafish embryos have been widely used in order to study heart formation, due to their transparency and its partial independence from the cardiovascular system. For vertebrates, the heart is the first organ that forms and starts operating [8]. Constrictions and bending (folding) are key elements in the early morphogenetic shap-

*Electronic address: luciano@if.sc.usp.br
†Electronic address: bant@if.sc.usp.br

ing of the heart tube. The spatial gene expression data considered in this work was acquired through the observation of 42-hour post-fertilization transgenic zebrafish embryos expressing EGFP specific for heart mesoderm myosin light chain (mlc2a-EGFP) [9]. The zebrafish embryos were anesthetized and kept fixed, and live-images of the heart were taken at ambient room temperature. The image recordings were made using a Nikon Eclipse TE300 inverted microscope using 20x/0.75 NA magnification. The microscope is coupled to a Bio-Rad Radiance MP 2100 scanning multiphoton confocal system (Cambridge,MA) with a two-photon Tsunami laser (Spectra Physics, CA). The GFP was excited with the two-photon laser, at 900 nm. The total dataset is composed of 110 confocal sections.

All the 110 confocal slices were combined so as to obtain the three-dimensional volume of the heart, from which the gene expression landscape was computed as described above. It is interesting to note that this scalar field can be visualized with direct volume rendering algorithms (DVR) [10]. In order to minimize the spatial quantization noise implied by digital image representation, gaussian smoothing was applied over the gene concentration data. This is done through the discrete convolution of a three-dimensional Gaussian kernel *k(x,y,z)* with the scalar field w, as expressed in Eq. (1)

$$\begin{aligned} w(x,y,z) * k(x,y,z) &= \sum_{i,j,k} w(i,j,k) \\ &\times \; k((x-i),(y-j),(z-k)) \end{aligned} \quad (1)$$

The smoothed reconstruction of the 3D gene activity pattern is shown in Figure 1a. The gradient of this scalar field was estimated by using the enhanced finite differences scheme described in [11], by convolving the gene expression concentration with three-dimensional masks. Next, we compute the lines of force by calculating the trajectories that maximize the gradient starting form arbitrary spatial positions sampled as points uniformly distributed through spheres centered at the three-dimensional volume.

The considered lines of force would correspond, for instance, to the putative path (set of 3D coordinates) followed by an object at position $\vec{r} = (x, y, z)$ with gradient dissipative dynamics:

$$\frac{\partial \vec{r}}{\partial t} = \vec{\nabla}\{\omega(x,y,z) * k(x,y,z)\}\,, \quad (2)$$

standard numerical integration was used in order to estimate such lines of force, which are illustrated in Figure 1b. The sampling criteria removed the lines whose scalar value of its end point were less than 10 (from a range of 0 to 255), eliminating those that do not reach the regions where mlc2a was being expressed. Small and too long trajectories were also removed, because they were influenced by noises. As expected, these lines converge to local maxima of the scalar gene expression field, which could be considered as *gene expression centers*. In analogy to graph theory, the total number of sampled lines of force converging to a specific center is referred to as the center *degree*. A total of 734 lines and 89 centers were obtained for the considered 3D gene expression data.

Figure 1b shows the sampled lines of force obtained by using the above described methodology, drawn in black or white according to thresholding criteria: the lines corresponding to gene expression activity centers with degrees smaller than 14 have been marked in white. Such threshold value was defined with basis on the relative frequency histogram of the distribution of centers degree, showed in Figure 2. It can be seen from Figure 1b that the genic activity centers exhibiting higher numbers of converging lines of force (marked black) tend to concentrate along the regions subjected to the constriction and folding implied by the heart formation dynamics (marked by arrow 1 in Figure 1a) as well as the sinus venosus (marked by arrow 2 in Figure 1a). The following biological interpretation are suggested in order to account for such result.

The heart forms from a tube of epimyocardial cells that express, among other genes, mlc2a. This gene is expressed uniformly throughout the heart, with the possibility of a weaker expression in the inflow pole, i.e. the region of the venous sinus and the atrium (Figure 1a). It is suggested here that the distribution of active cells could be determined by a gene activity field in such a way that the higher degree activity centers positively regulates the activation patterns of surrounding cells. The line of force pattern indicates that the expression of these cells coincides with morphogenetic events of heart formation, in particular the characteristic constrictions and bendings of the heart tube at the atrio-ventricular and the ventriculo-bulbar borders (arrows in Figure 1a), which are sites composed by high degree activity centers, involving cells more actively producing mlc2a. This process might be affected by the differential distribution of gene activation centers, as indicated by the respective numbers of lines of force which tended to be smoother at these locations.

While such hypotheses can only be verified through further experimental investigations, a novel methodology for 3D gene activity characterization has been shown to provide a natural and effective means for quantifying the spacial interactions between the biological structures involved in gene expression. Unlike differential measurements such as gradients or divergent magnitudes, the estimation of the lines of force and activity centers are integral features, indicating spatial interactions over substantial distances. It is expected that the proposed framework will prove to be useful in a number of other gene expression investigations, paving the way to a more objective understanding of the dynamics governing animal development and its pathologies.

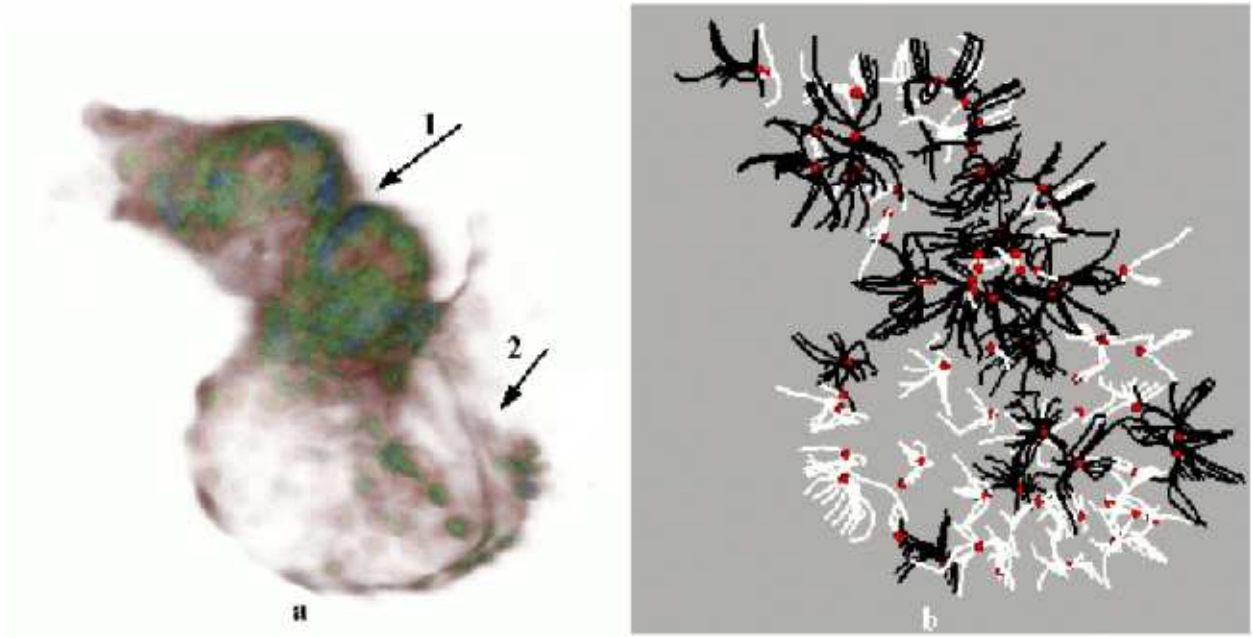

FIG. 1: (a) Visualization of the smoothed and reconstructed gene activity pattern of mlc2a during zebrafish heart formation. The inflow pole is on the upper left. Arrows 1 and 2 indicates the constriction/bending regions. The respective lines of force are shown in (b), segregated into black and white as described in the text.

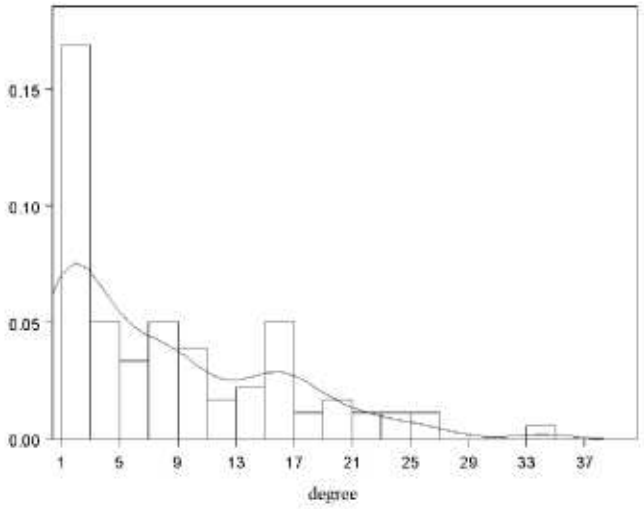

FIG. 2: Relative frequency histogram of the distribution of centers degree

Acknowledgments

The authors thank HFSP RGP39/2002 for funding this project. A. Azeredo are grateful to FAPESP (02/09149-2), and B. Travençolo is grateful to CAPES and FAPESP (03/13072-8) for financial support. M. Ibañes is partially supported by the Fulbright Program and Generalitat of Catalunya. Luciano da F. Costa is grateful to FAPESP (proc. 99/12765-2) and CNPq (proc. 301422/92-3) for financial support.

[1] S. B. Carroll, J. K. Grenier, and S. Weatherbee, *From DNA to diversity* (Blackwell Science: Malden, 2001).
[2] S. F. Gilbert, *Developmental Biology* (Sinauer Associates: Sunderland, 2003).
[3] J. Streicher, M. A. Donat, B. Strauss, R. Spörle, K. Schughart, and G. B. Müller, Nat. Rev. Genet. **25**, 147 (2000).
[4] Y. T. Tisen, Annu. Rev. Biochem. **67**, 3509 (1998).
[5] J. Lippincott-Schwartz and G. H. Patterson, Science **300**, 87 (2003).
[6] F. Schock and N. Perrimon, Annu. Rev. Cell. Dev. Biol. **18**, 463 (2002).
[7] S. Kuure, R. Vuolteenaho, and S. Vainio, Mech. Dev. **92**, 31 (2000).
[8] D. Y. R. Stainier, Nat. Rev. Genet. **2**, 39 (2001).
[9] A. Raya, C. Koth, D. Buscher, Y. Kawakami, T. Itoh, R. Raya, G. Sternik, H. Tsai, C. Rodriguez-Esteban, and J. Izpisua-Belmonte, Proc. Natl. Acad. Sci. **100**, 11889

(2003).

[10] W. J. Schröeder, K. Martin, and W. Lorensen, *The Visualization Toolkit - An Object Oriented Approach to 3D Graphics* (Prentice-hall, 1996), 3rd ed.

[11] S. W. Zucker and R. A. Hummel, IEEE Transactions on Pattern Analysis and Machine Intelligence **3(3)**, 324 (1981).